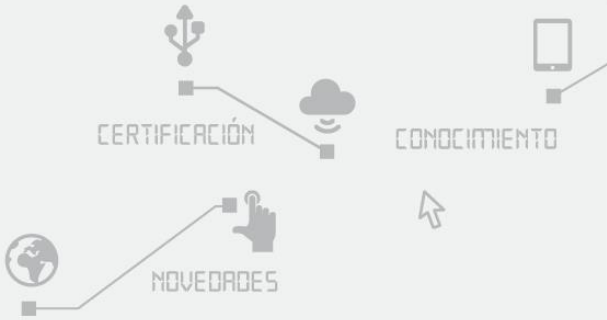
PUE DAY
ABRIL 2016



12:00 - 12:45

Cloudera Academy

“Enseñando tecnología Big Data en el aula”

Dan Johnson, Senior Education Manager EMEA at Cloudera
Ángel García, Product&Service Developer, PUE

cloudera®

pue IMPULSANDO EL CONOCIMIENTO TIC CUALIFICADO

Linux Professional Institute · cloudera · vmware IT ACADEMY · pue mobile CERTIFICATION · ORACLE ACADEMY · Microsoft Imagine Academy · Scrum Manager

cloudera®

Cloudera Academic Partnership

Teaching Hadoop to Tomorrow's Professionals

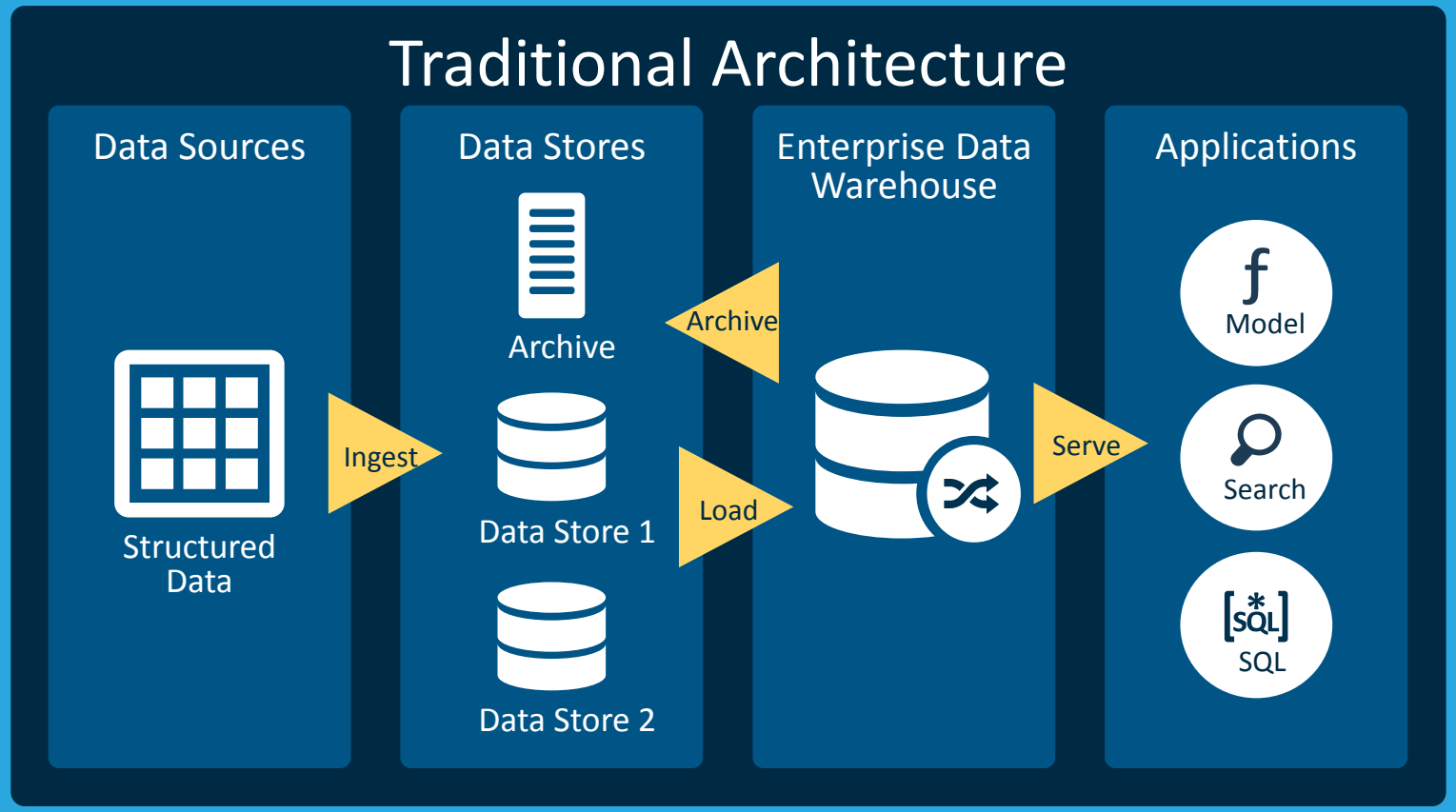
Big data problems & Hadoop solutions

cloudera

The Old Way

Complex, fragmented, costly

- Data siloed by department or line of business
- Mostly stored in expensive specialized systems
- Analysts only retrieve data into EDW when needed
- Nobody in the organization has a complete view

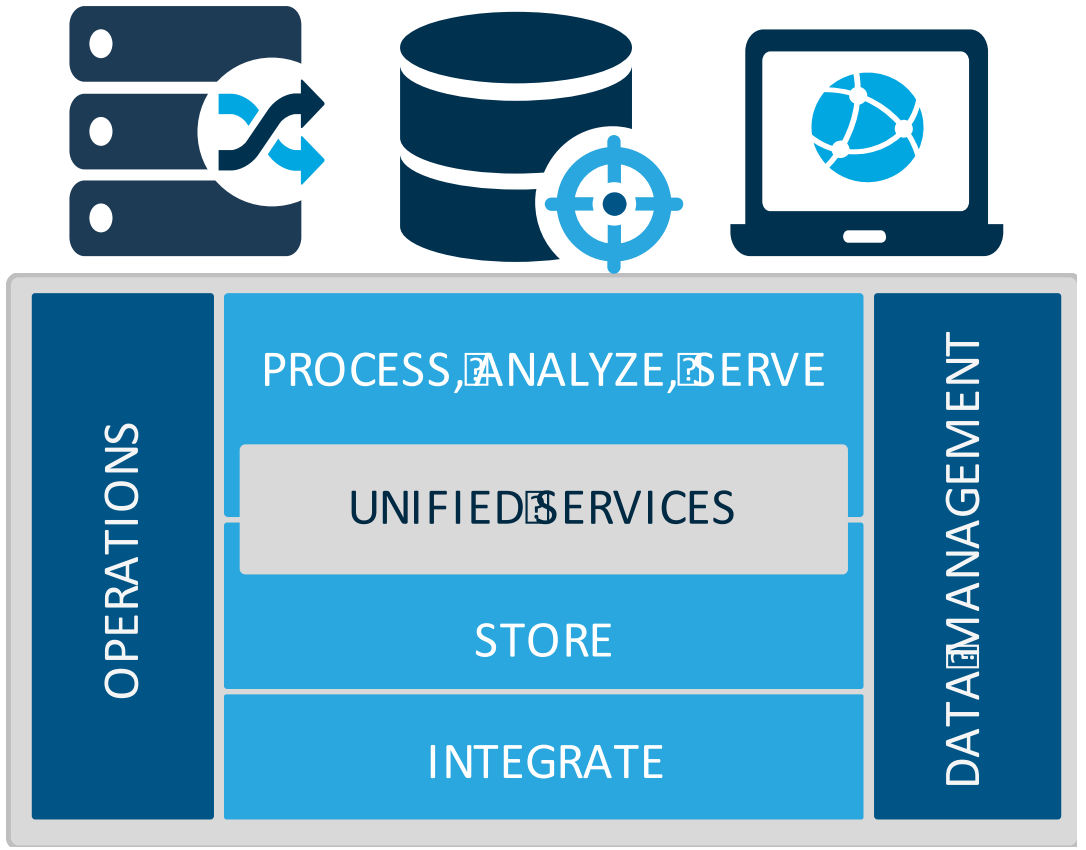


cloudera

The Hadoop Way

Simple Unified Efficient

- Data stored on scalable, low-cost platform
- Hadoop handles a variety of workloads
- Perform end-to-end workflows on a single system
- Provides broad data access across departments



cloudera

Hadoop Powers Big Data Breakthroughs

Delivering True Business Value through Customized Solutions

Financial Services & Banking

Banks ingest, transform, secure, and scrutinize data from multiple silos to detect/prevent fraud and mitigate risk

Healthcare & Biotech

Hospitals index and cross-query patient and pharma data in real time and manage next-gen sequencing workflows

Public Sector & Government

The intelligence community securely manages, audits, and analyzes massive video data sets to fight terrorism

Retail & E-Commerce

Take a 360º view of online behaviors to customize interactions, including real-time recommendations

Telecommunications

Mitigate against churn and get a full view of customers to provide targeted marketing and personalization

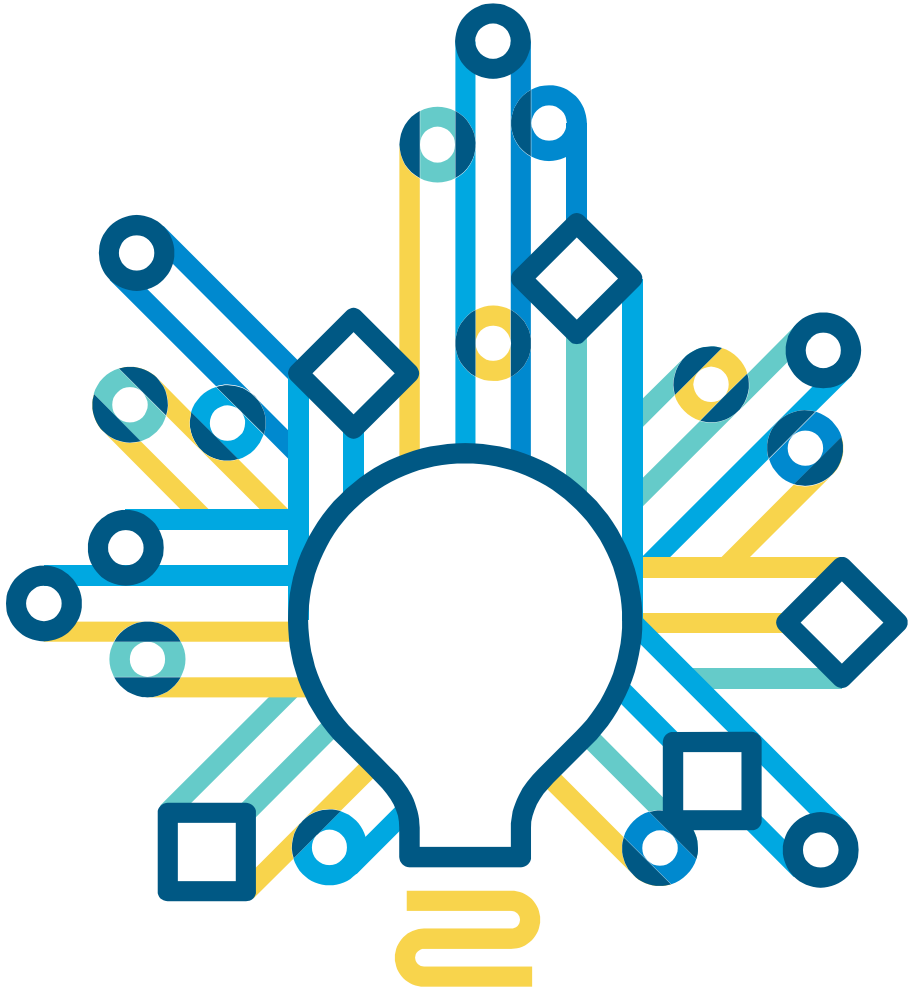
cloudera

The data skills gap

cloudera

Meet the need – big data talent shortage

- The highest paid skills are for Hadoop components or Cloudera, taking up **6 out of the top 10.**
- Hadoop skills are the highest paid in tech and are paid an **average of $124,927**
- Tech professional are more confident than ever, with **67% saying** that they will be able to find a new position



cloudera

Sources:
- http://media.dice.com/report/march-2015-big-data-markets/
- http://media.dice.com/report/check-and-check-u-s-technology-salaries-and-bonuses-rise/

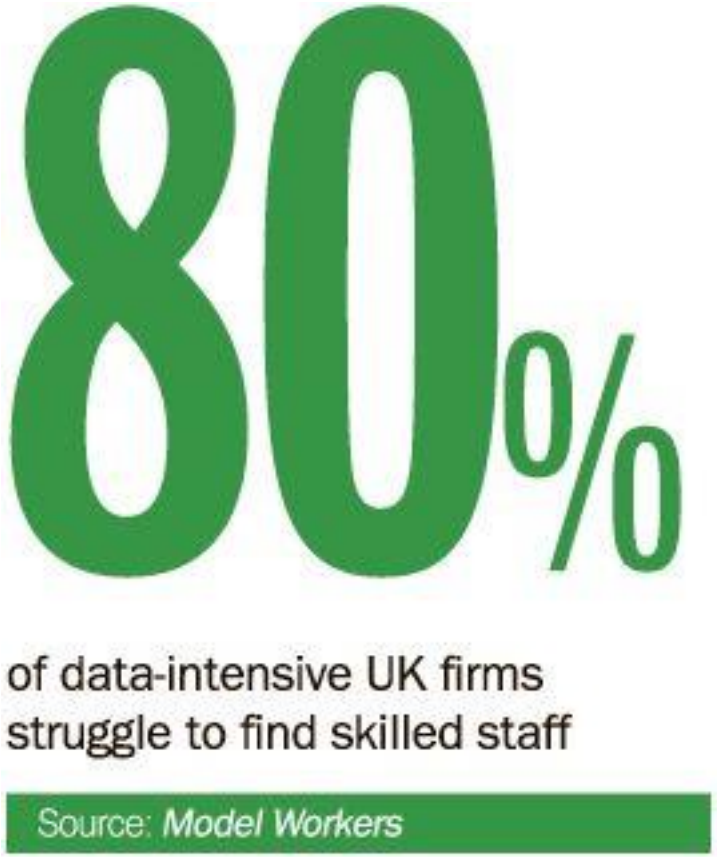
Global Talent Shortage

The United States, faces a shortage of

140-190k

people with analytical expertise

– McKinsey & Co

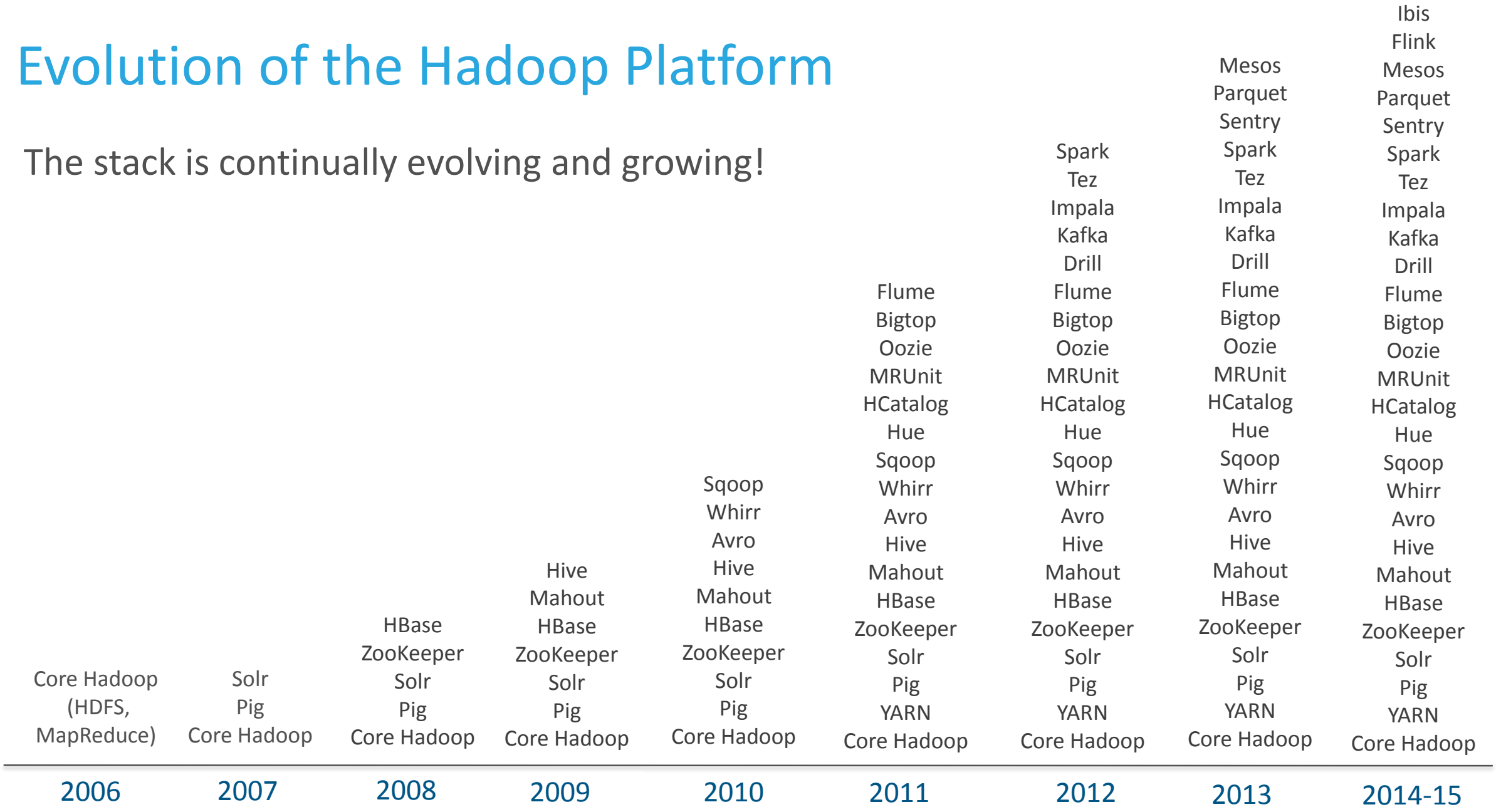


cloudera

Source: www.mckinsey.com/features/big_data
http://raconteur.net/technology/plugging-the-data-skills-gap

Evolution of the Hadoop Platform

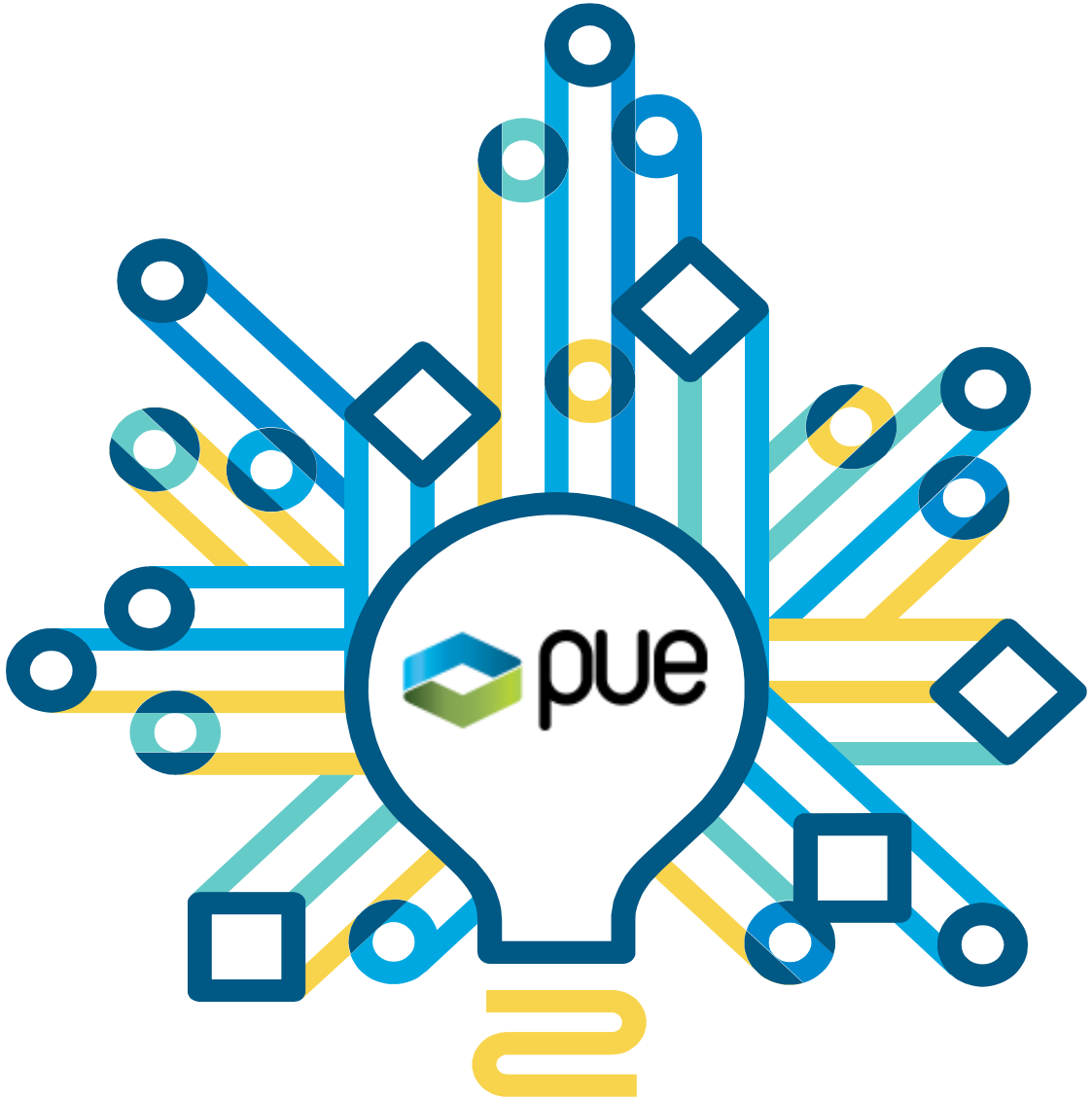
The stack is continually evolving and growing!



cloudera

PUE and Cloudera

- Authorised Cloudera Training Partner for more than 1 year
- Key representative of Cloudera in Spain
- **Leading our Cloudera Academic Partnership (CAP) in Spain for 2016**



cloudera

Cloudera Academic Program (CAP)

cloudera

Introducción

cloudera

La revolución Big Data que genera nuevas necesidades

El interés en Apache Hadoop y las soluciones Big Data ha conducido a una demanda creciente de profesionales con habilidades para generar valor a todo el volumen y tipos de datos que se producen.

Un 75% de las empresas invierten o planean hacerlo en Big Data

La tendencia de opinión es que los costes por almacenamiento y la mejora de la experiencia de los clientes son posibles gracias a la inversión en optimización de datos empresariales almacenados.

escrito por: Redacción Computing 18 de marzo 2016

Detrás del **Big Data** se oculta el perfil del cliente, sus hábitos de consumo, sus necesidades y las claves para llegar a ese tan deseado consumidor potencial, pero también se esconde **un 40% de información innecesaria**, que solo una correcta analítica de datos puede detectar, según señala **Wunderman**, compañía de customer intelligence, CRM, Real Time Marketing y marketing interactivo.

Dificultades para encontrar expertos en análisis de datos

Analytics Trends 2016 de Deloitte indica que la implantación de tecnologías cognitivas permitirá el desarrollo de nuevos puestos de trabajo.

escrito por: Redacción Computing 05 de febrero 2016

La creciente importancia que el **análisis de datos** está adquiriendo en todos los ámbitos de cualquier compañía está contribuyendo de forma significativa a la optimización y transformación de procesos. Sin embargo, la **falta de talento** capacitado para estas funciones es uno de los grandes retos a los que se enfrentarán en los próximos años. Así lo destaca el **informe Analytics Trends 2016**, elaborado por Deloitte.

más visto

cloudera | pue

¿Quién utiliza Hadoop?

cloudera | pue

¿Qué es Cloudera Academic Program?

cloudera

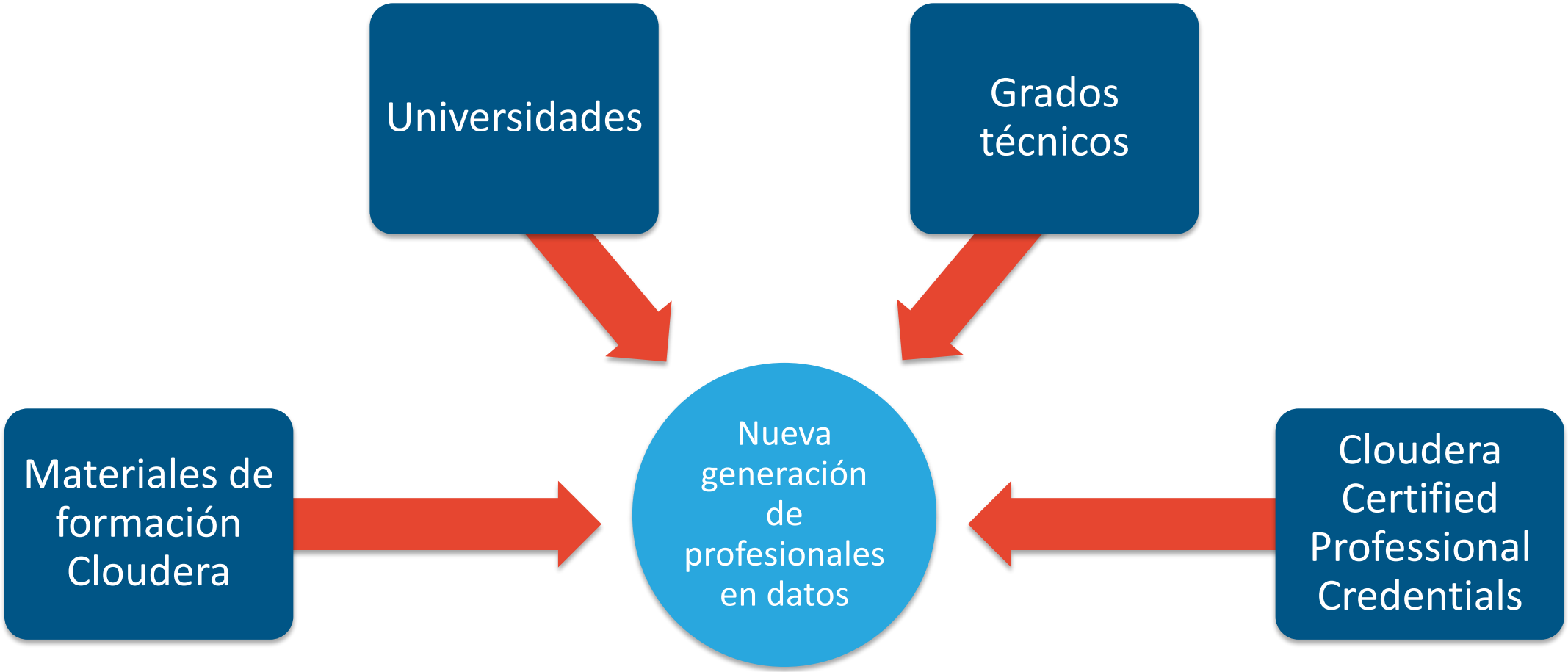
Cloudera Academic Program

- Programa académico de Cloudera que proporciona recursos y materiales para impartir formación oficial en Apache Hadoop
- CAP ha sido creado para ser una relación de colaboración que proporciona beneficios a todas las partes:

Estudiantes	Profesores	Instituciones académicas	Empresas	Cloudera
Acceso a los materiales de formación y certificación más populares de Hadoop, preparándoles para un trabajo en *data economy*	Guías de instructor, planes de estudio, máquinas virtuales preconfiguradas para ejercicios y descuentos en formaciones y certificaciones	Inclusión de formación Cloudera en el curriculum, proporcionando a los estudiantes habilidades diferenciadoras para encontrar trabajo	Un canal de talento desde instituciones académicas para ocupar necesidades de *internships* y empleabilidad	Introducir una nueva generación de usuarios de Hadoop, ayudando a asegurar que el mercado tiene el talento para ocupar el crecimiento de Big Data

cloudera | pue

Cloudera Academic Program



cloudera | pue

¿Qué ofrece Cloudera Academic Program?

cloudera

¿Qué ofrece Cloudera Academic Program?

Cloudera Academic Program

cloudera® ACADEMIC PARTNER

- Materiales curriculares
- Labs y VM
- Licencia Cloudera Manager
- Descuentos en cursos y certificaciones
- Descuentos en libros O'Reilly Media
- Materiales de Márketing

cloudera | pue

¿Qué ofrece Cloudera Academic Program?

Cloudera Academic Program

cloudera® ACADEMIC PARTNER

- Materiales curriculares
- Labs y VM
- Licencia Cloudera Manager
- Descuentos en cursos y certificaciones
- Descuentos en libros O'Reilly Media
- Materiales de Márketing

cloudera | pue

Cursos oficiales

1. Materiales curriculares para formación en Cloudera

Developing with Spark and Hadoop

Introduction to Hadoop and Big Data

cloudera | pue

Cursos oficiales

1. Materiales curriculares para formación en Cloudera

Developing with Spark and Hadoop

Introduction to Hadoop and Big Data

cloudera | pue

Cursos oficiales

Curso Developing with Spark and Hadoop

El curso cubre Spark y los elementos clave del ecosistema Hadoop utilizado en el desarrollo de aplicaciones finales para procesar eficientemente Big Data.

Una vez finalizado el curso, los alumnos entenderán los conceptos clave de Spark y Hadoop y aprenderán a aplicar herramientas e Spark y Hadoop en el desarrollo de aplicaciones

cloudera | pue

Cursos oficiales

Curso Developing with Spark and Hadoop

Duración: 36 horas (aprox.)

Prerrequisitos: Conocimiento en Scala o Python, línea de comandos Linux y SQL

Objetivos del curso:

- Encaje del ecosistema Hadoop con el ciclo de vida del procesamiento de datos
- Distribución, almacenamiento y procesado de datos en un cluster Hadoop
- Uso de Sqoop y Flume para ingerir datos
- Procesamiento de datos distribuidos con Spark
- Modelación de datos estructurado como tablas en Impala y Hive
- Elección del mejor formato de almacenamiento para diferentes patrones de uso de datos
- Mejores prácticas para el almacenamiento de datos

cloudera | pue

Cursos oficiales

1. Materiales curriculares para formación en Cloudera

Developing with Spark and Hadoop

Introduction to Hadoop and Big Data

cloudera | pue

Cursos oficiales

Curso Introduction to Hadoop and Big Data

El curso cubre la arquitectura Hadoop y el ecosistema de herramientas Hadoop. Estas tecnologías son la fundación del movimiento Big Data, que facilitan la gestión de la escalabilidad y el procesado de extensas cuantidades de datos.

cloudera | pue

Cursos oficiales

Curso Introduction to Hadoop and Big Data

Duración: 34 horas (aprox.)

Prerrequisitos: Conocimiento en Python, Java o C/C++, redes y sistemas y Linux

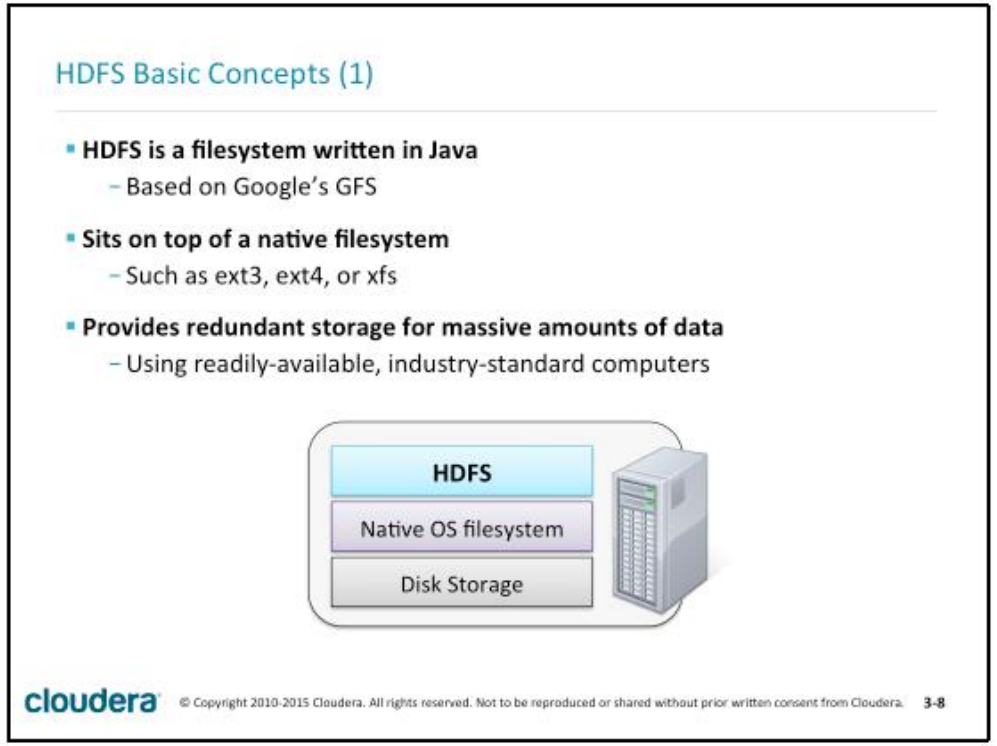
Objetivos del curso:

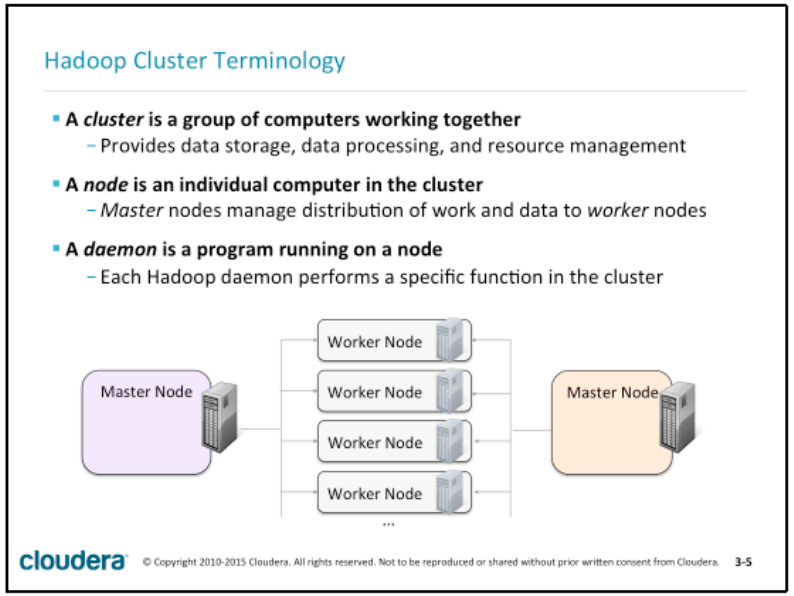
- Necesidad de utilizar Hadoop
- Conceptos del sistema de distribución de ficheros de Hadoop u MapReduce
- Resolución de problemas con Hadoop
- Principales tecnologías de Hadoop y el ecosistema Hadoop
- Desarrollar aplicaciones MapReduce
- Algoritmos frecuentes de MapReduce
- Uso de Pig y Hive para el desarrollo rápido de aplicaciones

cloudera | pue

Materiales curriculares

Guías del instructor





A cluster is a group of computers working together – the work is *distributed* across the cluster. As covered previously, distributed processing involves distributing both the tasks and the data the tasks operate on–and therefore a typical cluster has infrastructure for HDFS (to distribute the data) and whatever your chosen cluster management framework is (to distribute the processing).

cloudera | pue

Materiales curriculares

Slides para alumnos

What Is Apache Flume?

- **Apache Flume is a high-performance system for data collection**
  - Name derives from original use case of near-real time log data ingestion
  - Now widely used for collection of any streaming event data
  - Supports aggregating data from many sources into HDFS
- **Originally developed by Cloudera**
  - Donated to Apache Software Foundation in 2011
  - Became a top-level Apache project in 2012
  - Flume OG gave way to Flume NG (Next Generation)
- **Benefits of Flume**
  - Horizontally-scalable
  - Extensible
  - Reliable

Chapter Topics

Capturing Data with Apache Flume | **Introduction to Flume**

- What is Apache Flume?
- **Basic Flume Architecture**
- Flume Sources
- Flume Sinks
- Flume Channels
- Flume Configuration
- Conclusion
- Homework: Collect Web Server Logs with Flume

cloudera | pue

Materiales curriculares

Timings para clases y labs

Cloudera Academic Partnership

Course: CAP - Developing with Spark and Hadoop

Course Schedule with Timings

Week	Lecture Components	Component Duration (hours:min)	Class Duration (hours:min)	Comments	Homework Labs	Homework Duration (hours:min)
1	**Lecture 1 - Modules:** **Intorduction** **Introduction to Hadoop**					
	Course Overview and Logistics	0:30	0:30			
	1. Introduction	0:30	1:00			
	Break	0:05	1:05			
	2. Intro to Hadoop & Ecosystem (lecture part 1)	0:20	1:25			
	Demonstrate CAP Student VM Installation and Use	0:20	1:45	Topics to discuss: 1. How to download the CAP Student VM. 2. How to download and install VMWare. 3. How to run the VM. 4. Navigating within the VM, issuing Linux commands, issuing: *sudo jps*	Homework: 1. Install the CAP Student VM. 2. Follow instructions in Homework assignment guide for Chapter 2 - Setup and General Notes. 3. Verify that HDFS is working by issuing from the Linux command prompt: hdfs dfs -ls / 4. Read the Google GFS paper.	0:45
	Homework Discussion	0:20	2:05	Describe homework assignment due by next lecture.	*Optional homework: Textbook reading assignment*	

cloudera | pue

¿Qué ofrece Cloudera Academic Program?

Cloudera Academic Program

cloudera® ACADEMIC PARTNER

- Materiales curriculares
- Labs y VM
- Licencia Cloudera Manager
- Descuentos en cursos y certificaciones
- Descuentos en libros O'Reilly Media
- Materiales de Márketing

cloudera | pue

Labs y VM

Homework: Create and Populate Tables in Impala or Hive

Files and Directories Used in this Homework

Exercise directory: `$DEV1/exercises/impala`

MySQL Database: `loudacre`
MySQL Tables: `device`

Data files (HDFS): `/loudacre/webpage`

In these exercises you will define Impala/Hive tables to model and view data in HDFS.

Note: The accounts data will not be used in this exercise but will in a subsequent exercise.

You may perform this and subsequent exercises in either Impala or Hive. Most of the instructions are the same whichever tool you choose; where the instructions differ, the difference is noted. Following whichever set of instructions you prefer; if you have no preference, we suggest Impala because it is faster.

In this Exercise you will explore the performance effect of caching (that is, persisting to memory) an RDD.

1. Make sure the Spark Shell is running. If it isn't, restart it (in local mode with 2 threads) and paste in the job setup code from the previous exercise.
2. This time to start the job you are going to perform a slightly different action than last time: count the number of user accounts with a total hit count greater than five:

```
pyspark> accounthits\
  .filter(lambda (firstlast,hitcount): hitcount > 5)\
  .count()
```

```
scala> accounthits.filter(pair => pair._2 > 5).count()
```

3. Cache (persist to memory) the RDD by calling .
4. In your browser, view the Spark Application UI and select the **Storage** tab. At this point, you have marked your RDD to be persisted, but have not yet

cloudera | pue

¿Qué ofrece Cloudera Academic Program?

Cloudera Academic Program

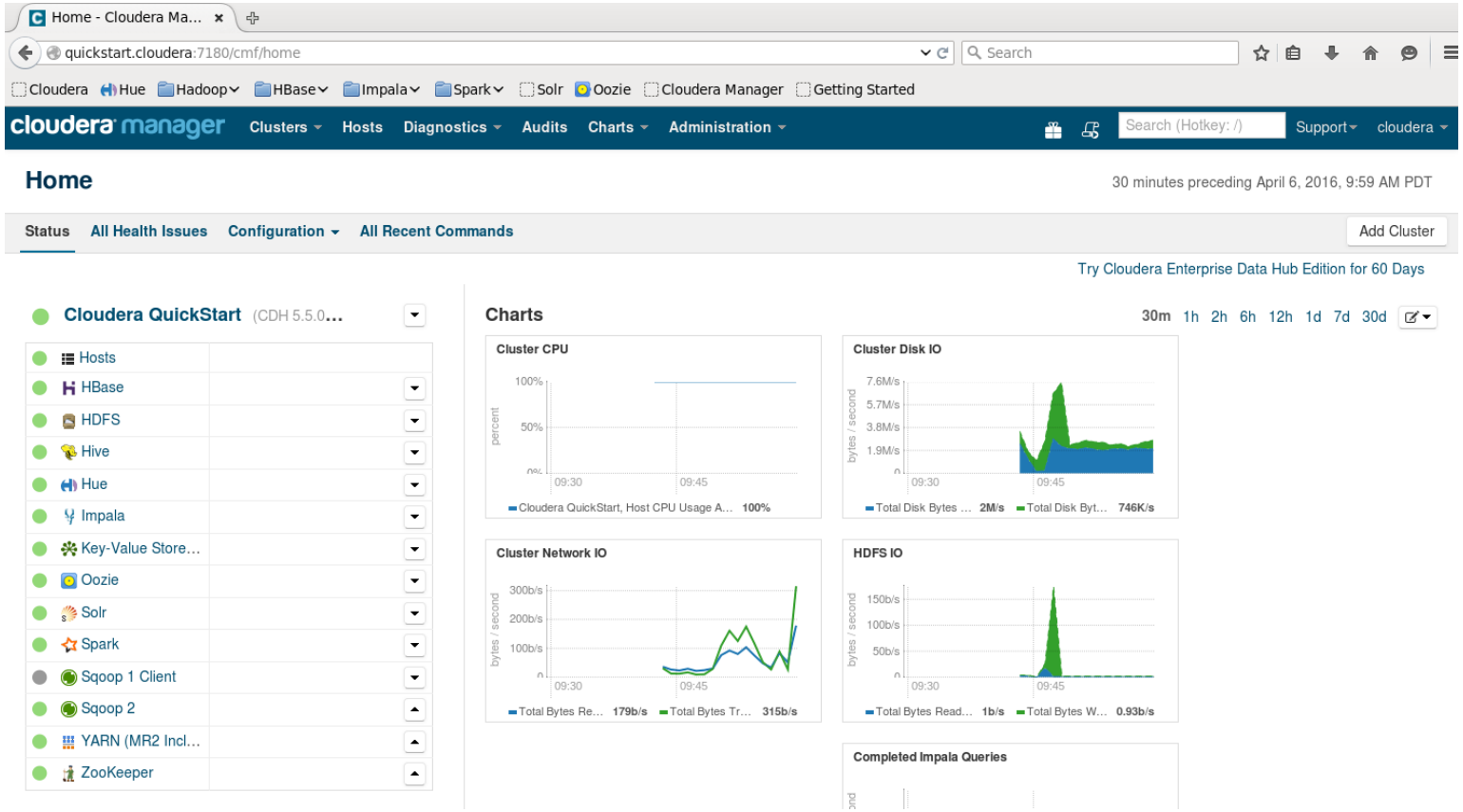
cloudera® ACADEMIC PARTNER

- Materiales curriculares
- Labs y VM
- Licencia Cloudera Manager
- Descuentos en cursos y certificaciones
- Descuentos en libros O'Reilly Media
- Materiales de Márketing

cloudera | pue

Licencia Cloudera Manager

- Licencia que permite todas las funcionalidades de Cloudera Manager
- No incluye servicios de soporte de Cloudera



¿Qué ofrece Cloudera Academic Program?

Cloudera Academic Program

cloudera® ACADEMIC PARTNER

- Materiales curriculares
- Labs y VM
- Licencia Cloudera Manager
- Descuentos en cursos y certificaciones
- Descuentos en libros O'Reilly Media
- Materiales de Márketing

cloudera | pue

Precios académicos para miembros CAP

- Descuentos en cursos privados y certificaciones Cloudera

Producto	Descuento aplicable para miembros de CAP
Cursos de calendario impartidos por Cloudera	50%
Cursos privados impartidos por Cloudera	30%
Certification Practice Exams	40%
Exámenes de certification	50%

cloudera | pue

¿Qué ofrece Cloudera Academic Program?

Cloudera Academic Program

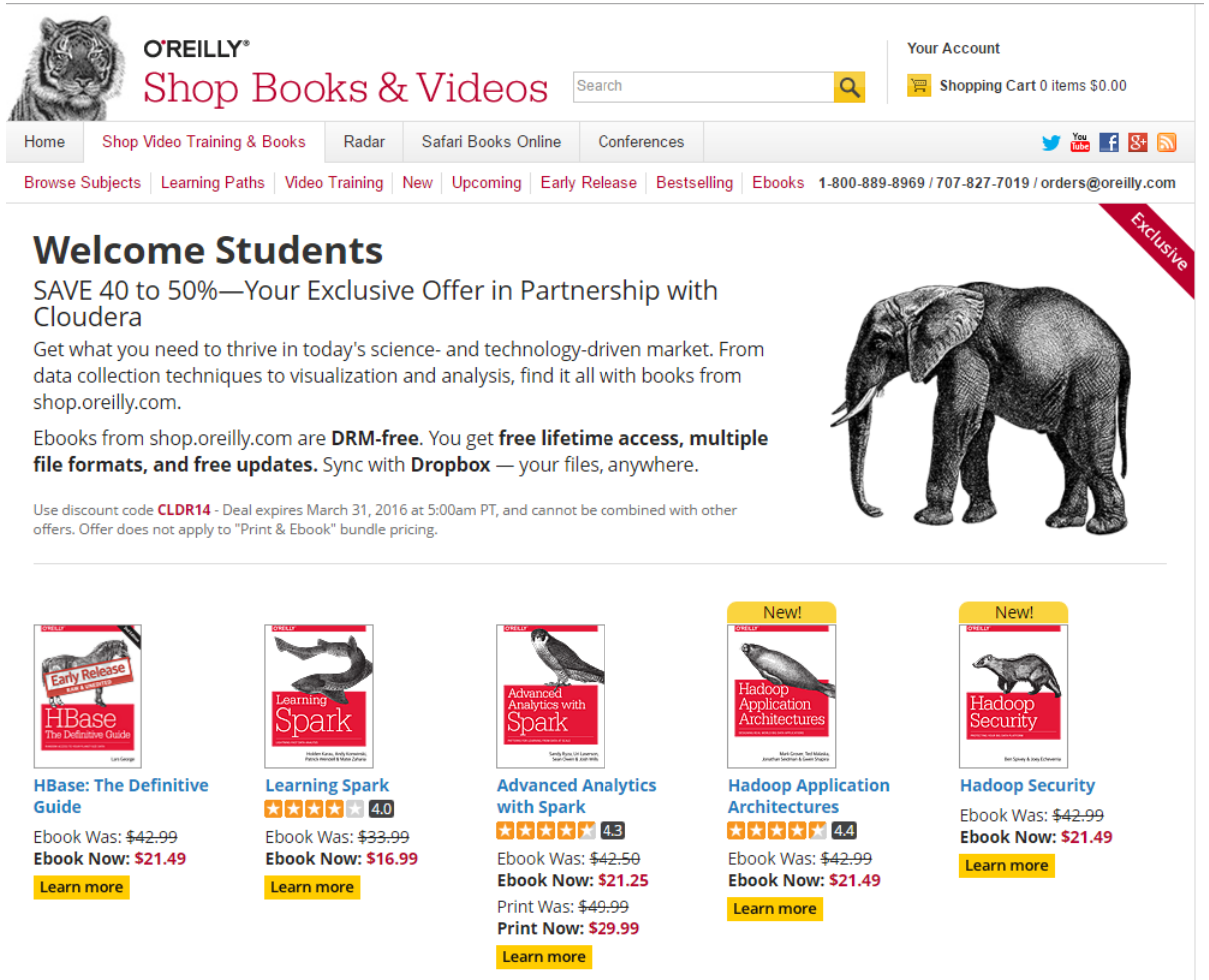
cloudera® ACADEMIC PARTNER

- Materiales curriculares
- Labs y VM
- Licencia Cloudera Manager
- Descuentos en cursos y certificaciones
- Descuentos en libros O'Reilly Media
- Materiales de Márketing

cloudera | pue

Descuentos libros O'Reilly Media

- Descuentos de hasta un 50% en la compra de libros distribuidos por O'Reilly Media
- 19 títulos disponibles relacionados con temáticas como Spark, Impala, Hive, Pig, etc.



¿Qué ofrece Cloudera Academic Program?

Cloudera Academic Program

cloudera® ACADEMIC PARTNER

- Materiales curriculares
- Labs y VM
- Licencia Cloudera Manager
- Descuentos en cursos y certificaciones
- Descuentos en libros O'Reilly Media
- Materiales de Márketing

cloudera | pue

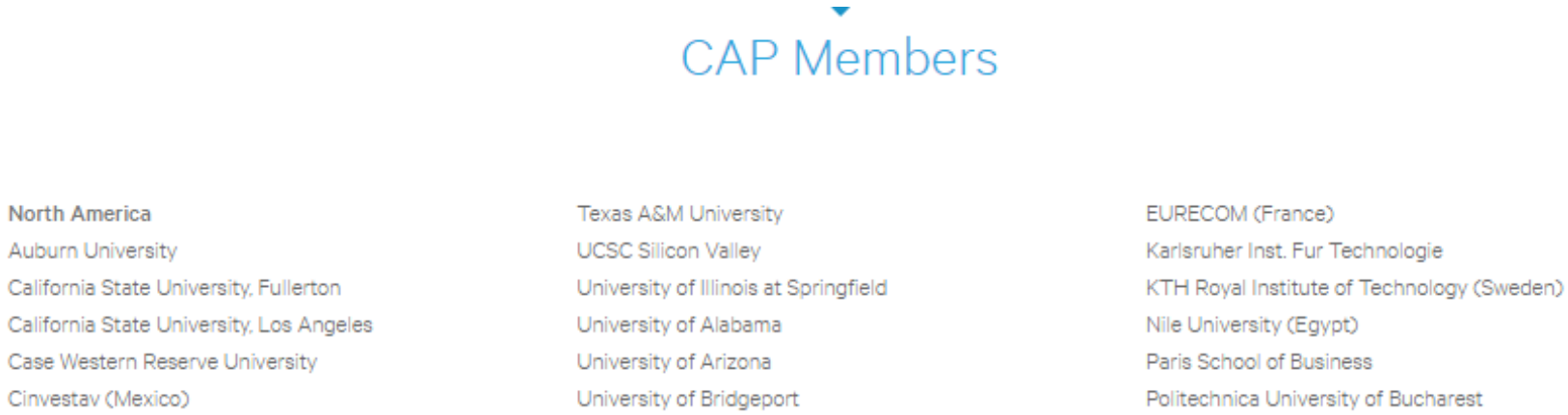
Materiales de Márketing

"Uso del logo oficial como miembro de Cloudera Academic Program"

cloudera | pue

Materiales de Márketing

“Aparición en el localizador de centros adheridos a Cloudera Academic Program”



http://www.cloudera.com/cap

cloudera | pue

Cloudera & PUE

cloudera

Cloudera & PUE

PUE ha establecido una relación de colaboración con Cloudera para la gestión e implementación de Cloudera Academic Program en España.

www.pue.es/cloudera-academy

cloudera | pue

Servicios de PUE

- Formación/videoconferencias (webinars) a los centros sobre la puesta en marcha y utilización de los recursos disponibles.
- Soporte, asesoramiento y resolución de consultas para obtener el mejor rendimiento del proyecto en función de su oferta educativa.
- Recomendaciones y ayuda en la puesta en marcha, implementación y utilización del programa.

cloudera | pue

Servicios de PUE

- Documentación específica sobre la utilización de los recursos para los centros educativos.
- Envío de información y explicación de las novedades del programa en castellano.
- Gestión directa con Cloudera en el desarrollo y mejora del programa.
- Velar por la correcta utilización del programa

cloudera | pue

Requisitos y suscripción

Requisitos de participación

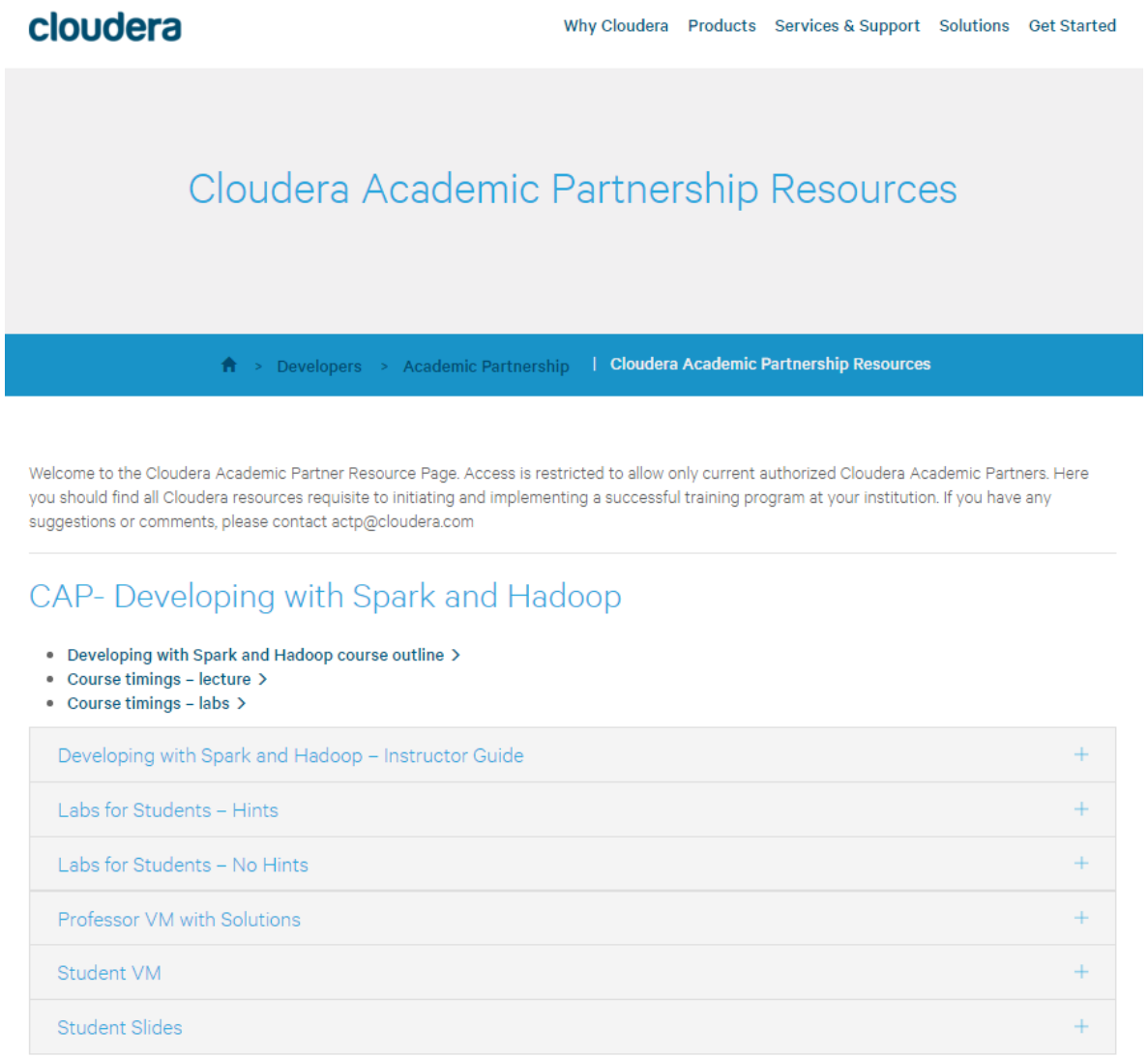
- Ser centro de formación reglada homologado para la impartición de Ciclos Formativos de Formación Profesional y Grados Universitarios.

- No aplicar los beneficios y recursos del programa fuera del ámbito de formación curricular reglada del centro

cloudera | pue

Suscripción

1. Cumplir los requisitos de incorporación al programa
2. Contactar con PUE
3. Abono de la cuota anual de 475,00€ (IVA Exento) en concepto de suscripción y soporte
4. Formalizar la adhesión a Cloudera como Cloudera Academic Partner
5. Recepción de claves de acceso



cloudera | pue

cloudera

Thank you

Àngel Garcia

Product & Service Developer at PUE

angel.garcia@pue.es

